Tillegg i pensum i MNFST101, våren 2003

Jarle Tufto

20. mars 2003

Dobbelforventing

Definisjon: La X og Y være to kontinuerlige stokastiske variabler definert over samme utfallsrom S. La $f_{X|y}(x)$ være betinget fordeling til X gitt y. Betinget forventning til X gitt y definerer vi da ved

$$E(X|y) = \int x f_{X|y}(x) dx. \tag{1}$$

På samme måte defineres en betinget varians ved

$$\mathrm{Var}(X|y) = \int (x-\mu)^2 f_{X|y}(x) dx. \tag{2}$$

Betinget forventning og varians til diskrete variable defineres på tilsvarende måte. Legg merke til at betingede forventninger og varianser generelt blir funksjoner av den variabelen vi betinger på.

Teorem: La X og Y være to kontinuerlige stokastiske variabler definert over samme utfallsrom S. Da er

$$E(X) = E_Y E(X|Y). \tag{3}$$

Ubetingede varianser kan uttrykkes ved betingede gjennom

$$\mathrm{Var}(X) = E_Y \,\mathrm{Var}(X|Y) + \mathrm{Var}_Y(E(X|Y)). \tag{4}$$

Bevis: Vi beviser bare (3). Vi innfører først funksjonen $\mu(y) = E(X|y)$. Da kan vi skrive den doble forventningen i (3) som

$$E_Y E(X|Y) = E(\mu(Y)). \tag{5}$$

Setning for forventning til en funksjon av en stokastisk variabel gir at dette er lik

$$\int \mu(y) f_Y(y) dy. \tag{6}$$

Setter vi inn definisjonen (1) av $\mu(y) = E(X|y)$ inn i (6) og bruker definisjonen av betinget sannsynlighetsfordeling blir uttrykket over lik

$$\begin{aligned}
&\int \left(\int x f_{X|y}(x) dx \right) f_Y(y) dy, \\
=&\int \int x f_{X|y}(x) f_Y(y) dx dy \\
=&\int \int x f_{X,Y}(x, y) dx dy \\
=&E(X).
\end{aligned} \tag{7}$$

Beviset for diskrete eller både diskrete og kontinuerlige variable kan gjøres på tilsvarende måte.

Blandinger

Lar vi parameteren i en fordeling være stokastisk og beskrevet ved en annen sannsynlighetsfordeling får vi det vi kaller en blanding ("mixture" på engelsk).

Anta at X gitt μ er Poissonfordelt med betinget fordeling

$$p_{X|\mu}(k) = \frac{\mu^k e^{-\mu}}{k!}, \tag{8}$$

og m.g.f.

$$M_{X|\mu}(t) = e^{\mu(e^t - 1)}. \tag{9}$$

Anta at parameteren μ i Poissonfordelingen over er gammafordelt med parametere r og λ. Da har μ m.g.f.

$$M_\mu(t) = \left(\frac{1}{1 - t/\lambda} \right)^r \tag{10}$$

Setning for dobbelforventing gir at den ubetingede forventningen til X er gitt ved

$$EX = E_\mu E(X|\mu) = E_\mu \mu = \frac{r}{\lambda}. \tag{11}$$

Tilsvarende setning for variansen (4) og uttrykk for forventning og varians i Poisson og gamma fordeling gir at

$$\begin{aligned}\operatorname{Var}(X) =& E_\mu \operatorname{Var}(X|\mu) + \operatorname{Var}_\mu(E(X|\mu)) \\ =& E\mu + \operatorname{Var}\mu \\ =& \frac{r}{\lambda} + \frac{r}{\lambda^2}\end{aligned} \tag{12}$$

Vi ønsker også å bestemme hele (den ubetingede) fordelingen til X,

$$p_X(k) = \int_0^\infty p_{X|\mu}(k) f_\mu(\mu) d\mu. \tag{13}$$

Dette er enklest å gjøre ved hjelp av moment genererende funksjoner.

Setning om dobbelforvetning og bruk av (9) gir at m.g.f. til X kan skrives som

$$\begin{aligned}M_X(t) &= E(e^{tX}) \\ &= E_\mu E(e^{tX}|\mu) \\ &= E_\mu M_{X|\mu}(t) \\ &= E_\mu e^{\mu(e^t-1)} \\ &= M_\mu(e^t-1),\end{aligned} \tag{14}$$

altså den momentgenererende funksjonen til μ innsatt $e^t - 1$. Setter vi $e^t - 1$ inn som argument i (10) får vi dermed, etter litt mellomregning, at

$$\begin{aligned}M_X(t) &= \left(\frac{1}{1-(e^t-1)/\lambda}\right)^r \\ &= \left(\frac{\lambda/(\lambda+1)}{1-(1-\lambda/(\lambda+1))e^t}\right)^r \\ &= \left(\frac{p}{1-(1-p)e^t}\right)^r\end{aligned} \tag{15}$$

for $p = \lambda/(\lambda+1)$. Fordi dette er m.g.f. til en negativ binomisk fordelt variabel med parametere p og r (definert slik at suksesser ikke er inkludert i opptellingen) er den ubetingede fordelingen til X som vi får når vi konstruerer blandingen over negativt binomisk fordelt.

Negativ binomisk fordeling kan altså genereres av flere ulike underliggende prosesser.

Regneoppgave

Anta at Y gitt σ^2 er $N(0, \sigma^2)$ og at σ^2 er eksponentiellt fordelt med parameter λ. Verifiser at den blandingen vi da får (den ubetingede fordelingen til Y) har fordeling på formen

$$f_Y(y) = \frac{1}{2}\gamma e^{-\gamma|y|}, -\infty < y < \infty, \tag{16}$$

ved å bestemme m.g.f. til blandingen og fordelingen over. Fordelingen over kalles Laplace eller dobbel eksponentiell fordeling.